

1 **Title:** The receptor binding domain of the new MERS coronavirus maps to a 231-residue region
2 in the spike protein that efficiently elicits neutralizing antibodies

3

4 **Running title:** Receptor binding domain of the MERS-CoV spike protein

5

6 **Authors:** Huihui Mou¹, V. Stalin Raj², Frank J.M. van Kuppeveld¹, Peter J.M. Rottier¹, Bart L.

7 Haagmans² and Berend Jan Bosch^{1,#}

8

9 **Affiliations:**

10 1) Department of Infectious Diseases and Immunology, Virology Division, Faculty of Veterinary

11 Medicine, Utrecht University, Yalelaan 1, 3584 CL, Utrecht, the Netherlands

12 2) Department of Viroscience, Erasmus Medical Center, Rotterdam, the Netherlands

13

14 # Corresponding author: b.j.bosch@uu.nl

15

16 **Abstract:**

17 The spike (S) protein of the recently emerged human coronavirus (MERS-CoV) mediates
 18 infection by binding to the cellular receptor dipeptidyl peptidase 4 (DPP4). Here we mapped the
 19 receptor-binding domain in the S protein to a 231-amino acid fragment (residues 358-588) by
 20 evaluating the interaction of spike truncation variants with receptor expressing cells and soluble
 21 DPP4. Antibodies to this domain - much less so to the preceding N-terminal region - efficiently
 22 neutralize MERS-CoV infection.

23

24 **Text**

25 Just 10 years following the outbreak of the severe respiratory acute syndrome coronavirus
 26 (SARS-CoV) the world is confronted with yet another deadly human coronavirus. The virus, first
 27 provisionally called human coronavirus-EMC (hCoV-EMC) (1, 2) but now named MERS-CoV,
 28 referring to its emergence in the Middle-East and to the respiratory syndrome it causes,
 29 belongs to the *betacoronavirus* genus lineage 2c (3). As of June 7th 2013, 55 cases have been
 30 laboratory confirmed including 31 deaths, all from - or linked to - the Arabian Peninsula (4). Like
 31 with SARS-CoV, patients affected by MERS-CoV suffer from severe and often lethal lower
 32 respiratory tract infection. The epidemiology of MERS-CoV is still enigmatic, but the
 33 geographical distribution of epidemiologically unlinked individuals points to intermittent,
 34 zoonotic transmission from a – so far unknown - animal source, whereas a number of reported
 35 clusters indicate limited human-to-human spread (5).

36 The main determinant of coronavirus tropism is the viral spike (S) protein as it mediates
 37 binding to a cell-surface receptor. The MERS-CoV S protein, a 1353 amino acid type I

38 membrane glycoprotein, assembles into trimers that constitute the spikes or peplomers on the
 39 surface of the enveloped coronavirus particle. The protein combines the two essential entry
 40 functions, namely that of host receptor binding and membrane fusion, which are attributed to
 41 the N-terminal (S1, residues 1-751) and C-terminal (S2, residues 752-1353) half of the S
 42 protein, respectively (Fig.1a). Recently we have identified dipeptidyl peptidase 4 (DPP4, also
 43 known as CD26), expressed in the human lung, as a functional receptor for MERS-CoV(6).
 44 Importantly, MERS-CoV can also use the evolutionary conserved DPP4 of other species, most
 45 notably that of bats(6, 7).

46 Coronaviruses bind to receptors via independently folded, generally about 150-300
 47 residues long receptor binding domains (RBD) present in their S1 subunit, of which the location
 48 within S1 can vary (8-10). Thus, for the *betacoronavirus* mouse hepatitis virus (MHV) the
 49 binding to its CEACAM receptor (11) has been mapped to the N-terminal ~300 amino acids of
 50 the spike protein (12, 13) whereas for the SARS-CoV - of the same genus - binding to the ACE2
 51 receptor (14) maps to residues 323-502 of S1 (15, 16) (Fig.1a). Identification of the RBD can
 52 hence help the development of monoclonal antibodies and vaccines for the treatment and
 53 prevention of infection. The RBD is the most important target for neutralizing antibodies (12, 17,
 54 18) preventing virus-receptor interaction.

55 We previously used the S1 domain of MERS-CoV fused to the Fc-region of human IgG to
 56 demonstrate the interaction of S1 with DPP4-expressing cells and with soluble, i.e. non
 57 membrane-anchored DPP4 (6). To identify the receptor binding domain in the MERS-CoV S1
 58 subunit, we generated S1-Fc protein chimera's with truncations at the C-terminus and
 59 N-terminus of the S1 domain. We considered a three domain structure of the MERS-CoV S1

60 protein (residues 1-357, 358-588 and 589-747) based on the predicted location and structure of
 61 the RBD of two other *betacoronaviruses*, MHV and SARS-CoV (12, 13, 15, 16), of which the
 62 homologous regions for MERS-CoV S map to the residues 18-351 and 379-580, respectively
 63 (Fig.1b). In addition, a soluble form of human DPP4 (residues 39-766) was made, which was
 64 C-terminally tagged with the Fc region. These proteins were expressed in HEK-293T cells after
 65 transfection of the respective expression plasmids and subsequently affinity-purified from the
 66 cell culture supernatant using protein A sepharose beads as described(6). The Fc region of
 67 purified sDPP4-Fc was proteolytically removed using trypsin (data not shown). First, we
 68 analyzed the S1-Fc proteins and C-terminal S1 truncations thereof for their ability to interact
 69 with sDPP4 using a co-purification assay. sDPP4 was efficiently co-purified by the S1-Fc
 70 variants encompassing residues 1-588 and 1-747 whereas the 1-357 S1-Fc variant was unable
 71 to bind sDPP4 (Fig.2a). We next generated an S1-Fc variant comprising residues 358-588, a
 72 region homologous to the ACE2 receptor binding domain in SARS-CoV S1 (Fig.2a). This S1-Fc
 73 truncation variant efficiently bound soluble DPP4, indicating that the DPP4 receptor binding
 74 domain is located within the 358-588 residues domain of the MERS-CoV spike protein.

75 We subsequently tested the ability of these S1-Fc variants to bind to HEK-293T cells
 76 transiently expressing DPP4 by using flow cytometry. The S1-Fc variants encompassing
 77 residues 1-588 and 358-588 bound to DPP4-expressing HEK-293T cells with efficiencies
 78 comparable to the full-length S1 protein whereas no binding was observed with the 1-357
 79 S1-Fc variant (Fig.2b). These data show the 358-588 amino acids S1 region to be essential and
 80 sufficient for binding to DPP4-expressing cells, consistent with the results of the sDPP4
 81 interaction study.

82 To confirm the observed interactions in a more biological assay we analyzed the ability of
83 the S1-Fc variants to prevent MERS-CoV infection. Thus, Huh-7 cells were preincubated with
84 the different S1-Fc variants before being inoculated with the MERS-CoV. We found that the
85 variants encompassing residues 1-747 and 358-588, but not the 1-357 S1-Fc variant, inhibited
86 infection (Fig.2c).

87 Finally polyclonal antibodies were raised in rabbits against the 1-747, 1-357 and 358-588
88 S1-Fc variants (Davids Biotechnology GmbH, Germany). The sera, which displayed equal
89 ELISA titers towards its antigen (1:300.000, data not shown), were tested for their ability to
90 neutralize virus infectivity. Antibodies raised against the 358-588 S1-Fc variant efficiently
91 neutralized virus infectivity, superior to those raised against the 1-747 and 1-357 S1-Fc variants
92 (Fig.2d). This indicates that neutralizing epitopes within S1 are primarily localized to the RBD
93 region. The elicited antibodies are likely to block the interaction of the spike protein with DPP4
94 thereby neutralizing MERS-CoV infectivity. Of note, antibodies raised against the MERS-CoV-S
95 RBD did not cross-neutralize SARS-CoV infection (data not shown). The results demonstrate the
96 potential of S1 protein and of the 358-588 S1 polypeptide as subunit vaccines with a high
97 biosafety profile compared to vaccines based on inactivated viruses or live-attenuated virus.

98 Except for the *betacoronavirus* MHV, which binds to its CEACAM receptor through a
99 domain in the N-terminal part of its S1 protein, the RBDs of all other coronaviruses that engage
100 protein receptors and that have been mapped occur in the C-terminal portion of the S1 subunit
101 (Fig.3). Examples also include the *alphacoronaviruses* binding to ACE2 (hCoV-NL63) and APN
102 (e.g. TGEV, hCoV-229E) (10, 19-25). In this study we have experimentally mapped the RBD of
103 MERS-CoV to a 231-amino acid fragment (residues 358-588) within the spike protein. This

domain nicely corresponds with the S1 region recently anticipated to interact with the DPP4 receptor on the basis of theoretical S1 structure predictions (26). The RBD in the MERS-CoV S1 protein localizes in the same region where the SARS-CoV S protein interacts with its ACE2 receptor (26). The SARS-CoV RBD structure displays a 5-stranded β -sheet core structure (β 1-4 and β 7) maintaining the overall domain conformation, and a long extended loop containing two anti-parallel β -sheets (β 5 and β 6) responsible for receptor binding(16). Intriguingly, compared to SARS-CoV, the RBD of MERS-CoV contains a relatively conserved core domain but a highly variable loop region, tentatively explaining the differential receptor usage(26). Crystallization and structure analysis of this MERS-CoV RBD region in complex with DPP4 will give detailed insight into the virus-receptor binding interface.

114

115 **Figure legends**

116 **Figure 1.** Receptor binding domains in *betacoronavirus* spike proteins and S1-Fc expression constructs. a, Schematic representation of the *betacoronaviruses* SARS-CoV, hCoV-EMC S and MHV (strain A59) spike (S) protein sequence (drawn to scale) aligned at the S1-S2 junction. 117
118 The known receptor binding domain in the S1 subunit of MHV and SARS-CoV S proteins and their corresponding homologous regions in hCoV-EMC S as defined by ClustalW alignment are 119
120 indicated. Positions of the transmembrane domain (yellow bar; predicted by the [TMHMM](#) server) and of the predicted *N*-glycosylation sites (Ψ ; predicted by the [NetNGlyc](#) server, only shown for 121
122 the hCoV-EMC S) are indicated. The border between the S1 and S2 subunits of the spike protein is represented by a vertical white line. b, upper panel, schematic presentation of the 123
124 hCoV-EMC S1 subunit (residues 1-751) sequence. Cysteine positions in S1 subunit are 125

126 indicated by vertical white lines with corresponding amino acid positions on top. Positions of
 127 cysteines highly conserved among betacoronaviruses S1 proteins are in bold. Predicted
 128 disulfide bond connections inferred from the structure of the SARS-CoV receptor binding
 129 domain are presented as connecting black lines underneath. Lower panel, domains of the
 130 hCoV-EMC S1 subunit expressed as Fc chimera's.

131

132 **Figure 2.** The DPP4 binding domain is located within residues 358-588 of the hCoV-EMC spike
 133 protein and efficiently elicits neutralizing antibodies. **a**, S1-Fc chimeric proteins and soluble
 134 DPP4 (sDPP4) receptor were expressed from HEK-293T cells and purified from the culture
 135 supernatant. S1-Fc proteins were mixed with sDPP4 followed by protein A sepharose affinity
 136 isolation, analyzed on a NoVEX® 4-12% Tris-Glycine gradient gel under non-reducing
 137 conditions, and stained with GelCodeBlue reagent. Position of the S1-Fc proteins - running as
 138 dimers under non-reducing conditions due to an Fc interchain disulphide bond - and sDPP4 as
 139 well as the sizes of the marker proteins are indicated. Individual proteins were loaded as
 140 controls. **b**, Binding of hCoV-EMC S1-Fc proteins to DPP4 expressing cells. 2.5×10^5 HEK-293T
 141 cells transfected with control pCAGGS (grey shaded area) or with pCAGGS-DPP4 (black line)
 142 expression plasmid were incubated with 15 µg/ml of the indicated S1-Fc followed by incubation
 143 with DyLight488 labeled goat-anti-human IgG antibody and analysis by flow cytometry. An
 144 Fc-chimera containing the S1 of infectious bronchitis virus (IBV-S1-Fc) was taken along as a
 145 negative control. **c**, Inhibition of hCoV-EMC infection by S1-Fc 1-747, 1-357 and 358-588
 146 variants. Huh7 cells were preincubated with 40 µg/ml S1-Fc 1-747, 1-357 or 358-588 for 0.5 h
 147 prior to virus inoculation (1 h), all at RT. Mock incubated cells (control) and cells incubated with

148 a DPP4 polyclonal antibody (anti-DPP4) were taken along as controls. Following incubation at
 149 37° C for 8 hours, infected cells were detected by immunofluorescence and infection was
 150 quantified (relative to control). The experiment was carried out twice and the data of one a
 151 representative experiment are shown. Error bars indicate standard error of the mean. **d**,
 152 Neutralization of hCoV-EMC infection by rabbit antisera raised against the S1-Fc 1-747, 1-357
 153 and 358-588 variants. Virus (200 pfu) was premixed 1:1 with serial dilutions of sera obtained
 154 (open bars) or after immunization (closed bars) prior to inoculation onto VERO cells and virus
 155 infection was monitored by the occurrence of CPE at 72 hours post infection. Virus
 156 neutralization titers (VNT) were determined in quadruplicate as the highest serum dilutions that
 157 completely prevent CPE. The experiment was carried out twice and the data of one a
 158 representative experiment are shown. Error bars indicate standard error of the mean.

159

160 **Figure 3.** Localization of receptor-binding domains in coronavirus spike proteins. Schematic
 161 presentation of the spike proteins of the *alphacoronaviruses* TGEV and hCoV-NL63 and of the
 162 *betacoronaviruses* SARS-CoV, hCoV-EMC and MHV (drawn to scale), aligned at the S1-S2
 163 junction. Blue boxes represent the receptor-binding domains (RBD) and indicate the engaged
 164 receptor. The RBD of TGEV, hCoV-NL63, SARS-CoV and MHV have been confirmed by
 165 crystallography (12, 15, 22, 26). Grey boxes indicate the transmembrane domain. Sequence
 166 IDs: TGEV (ABG89335.1), hCoV-NL63 (NC_005831.2), SARS-CoV (NP_828851.1),
 167 hCoV-EMC (AFS88936.1), MHV (NC_001846.1).

168

Acknowledgements We thank Ger Arkesteijn and Laura de Vries (UU, Utrecht, The Netherlands) for experimental support. This work was supported by a fellowship from China Scholarship Council to H.M. The study was financed by the European Union FP7 projects EMPERIE (contract number 223498) and ANTIGONE (contract number 278976).

References

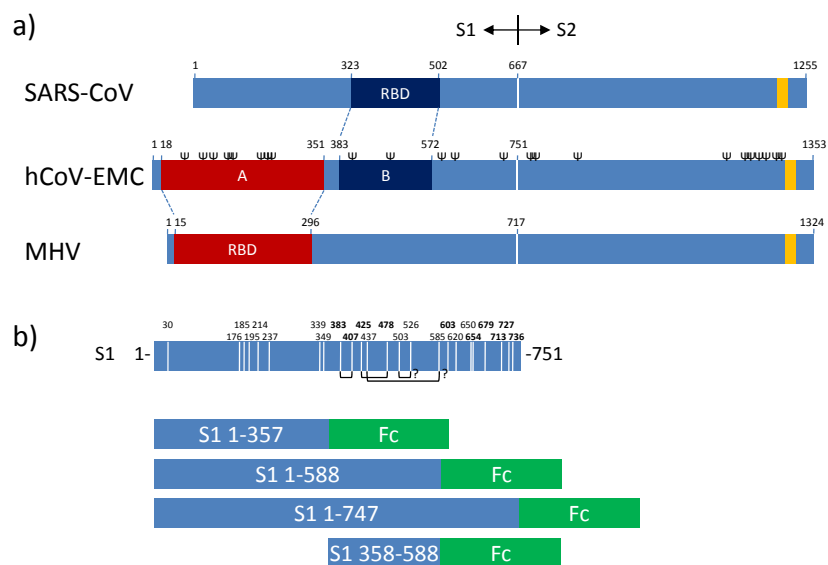
1. Zaki, A. M., S. van Boheemen, T. M. Bestebroer, A. D. Osterhaus, and R. A. Fouchier. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**:1814-1820. doi: 10.1056/NEJMoa1211721; 10.1056/NEJMoa1211721.
2. van Boheemen, S., M. de Graaf, C. Lauber, T. M. Bestebroer, V. S. Raj, A. M. Zaki, A. D. Osterhaus, B. L. Haagmans, A. E. Gorbalenya, E. J. Snijder, and R. A. Fouchier. 2012. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio.* **3**:10.1128/mBio.00473-12. doi: 10.1128/mBio.00473-12; 10.1128/mBio.00473-12.
3. de Groot, R. J., S. C. Baker, R. S. Baric, C. S. Brown, C. Drosten, L. Enjuanes, R. A. Fouchier, M. Galiano, A. E. Gorbalenya, Z. Memish, S. Perlman, L. L. Poon, E. J. Snijder, G. M. Stephens, P. C. Woo, A. M. Zaki, M. Zambon, and J. Ziebuhr. 2013. Middle East Respiratory Syndrome Coronavirus (MERS-CoV); Announcement of the Coronavirus Study Group. *J. Virol.* . doi: 10.1128/JVI.01244-13.
4. Source WHO. http://www.who.int/csr/don/2013_06_07/en/index.html.
5. Health Protection Agency (HPA) UK Novel Coronavirus Investigation team. 2013. Evidence of person-to-person transmission within a family cluster of novel coronavirus infections, United Kingdom, February 2013. *Euro Surveill.* **18**:20427.
6. Raj, V. S., H. Mou, S. L. Smits, D. H. Dekkers, M. A. Muller, R. Dijkman, D. Muth, J. A. Demmers, A. Zaki, R. A. Fouchier, V. Thiel, C. Drosten, P. J. Rottier, A. D. Osterhaus, B. J. Bosch, and B. L. Haagmans. 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature.* **495**:251-254. doi: 10.1038/nature12005; 10.1038/nature12005.
7. Muller, M. A., V. S. Raj, D. Muth, B. Meyer, S. Kallies, S. L. Smits, R. Wollny, T. M. Bestebroer, S. Specht, T. Suliman, K. Zimmermann, T. Binger, I. Eckerle, M. Tschapka, A. M. Zaki, A. D. Osterhaus, R. A. Fouchier, B. L. Haagmans, and C. Drosten. 2012. Human coronavirus EMC does not require the SARS-coronavirus receptor and maintains broad

- 201 replicative capability in mammalian cell lines. MBio. **3**:10.1128/mBio.00515-12. doi:
202 10.1128/mBio.00515-12; 10.1128/mBio.00515-12.
- 203 **8. Graham, R. L., and R. S. Baric.** 2010. Recombination, reservoirs, and the modular spike:
204 mechanisms of coronavirus cross-species transmission. J. Virol. **84**:3134-3146. doi:
205 10.1128/JVI.01394-09; 10.1128/JVI.01394-09.
- 206 **9. Li, W., S. K. Wong, F. Li, J. H. Kuhn, I. C. Huang, H. Choe, and M. Farzan.** 2006. Animal
207 origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein
208 interactions. J. Virol. **80**:4211-4219. doi: 10.1128/JVI.80.9.4211-4219.2006.
- 209 **10. Li, F.** 2012. Evidence for a common evolutionary origin of coronavirus spike protein
210 receptor-binding subunits. J. Virol. **86**:2856-2858. doi: 10.1128/JVI.06882-11;
211 10.1128/JVI.06882-11.
- 212 **11. Williams, R. K., G. S. Jiang, and K. V. Holmes.** 1991. Receptor for mouse hepatitis virus
213 is a member of the carcinoembryonic antigen family of glycoproteins. Proc. Natl. Acad. Sci. U.
214 S. A. **88**:5533-5536.
- 215 **12. Kubo, H., Y. K. Yamada, and F. Taguchi.** 1994. Localization of neutralizing epitopes and
216 the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus
217 spike protein. J. Virol. **68**:5403-5410.
- 218 **13. Peng, G., D. Sun, K. R. Rajashankar, Z. Qian, K. V. Holmes, and F. Li.** 2011. Crystal
219 structure of mouse coronavirus receptor-binding domain complexed with its murine receptor.
220 Proc. Natl. Acad. Sci. U. S. A. **108**:10696-10701. doi: 10.1073/pnas.1104306108;
221 10.1073/pnas.1104306108.
- 222 **14. Li, W., M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J.**
223 **L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan.** 2003.
224 Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature.
225 **426**:450-454. doi: 10.1038/nature02145.
- 226 **15. Wong, S. K., W. Li, M. J. Moore, H. Choe, and M. Farzan.** 2004. A 193-amino acid
227 fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2.
228 J. Biol. Chem. **279**:3197-3201. doi: 10.1074/jbc.C300520200.
- 229 **16. Li, F., W. Li, M. Farzan, and S. C. Harrison.** 2005. Structure of SARS coronavirus spike
230 receptor-binding domain complexed with receptor. Science. **309**:1864-1868. doi:
231 10.1126/science.1116480.
- 232 **17. Bonavia, A., B. D. Zelus, D. E. Wentworth, P. J. Talbot, and K. V. Holmes.** 2003.
233 Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus
234 HCoV-229E. J. Virol. **77**:2530-2538.

- 235 18. **He, Y., Y. Zhou, S. Liu, Z. Kou, W. Li, M. Farzan, and S. Jiang.** 2004. Receptor-binding
236 domain of SARS-CoV spike protein induces highly potent neutralizing antibodies: implication
237 for developing subunit vaccine. *Biochem. Biophys. Res. Commun.* **324**:773-781. doi:
238 10.1016/j.bbrc.2004.09.106.
- 239 19. **Breslin, J. J., I. Mork, M. K. Smith, L. K. Vogel, E. M. Hemmila, A. Bonavia, P. J.**
240 **Talbot, H. Sjostrom, O. Noren, and K. V. Holmes.** 2003. Human coronavirus 229E: receptor
241 binding domain and neutralization by soluble receptor at 37 degrees C. *J. Virol.* **77**:4435-4438.
- 242 20. **Hofmann, H., G. Simmons, A. J. Rennekamp, C. Chaipan, T. Gramberg, E. Heck, M.**
243 **Geier, A. Wegele, A. Marzi, P. Bates, and S. Pohlmann.** 2006. Highly conserved regions
244 within the spike proteins of human coronaviruses 229E and NL63 determine recognition of their
245 respective cellular receptors. *J. Virol.* **80**:8639-8652. doi: 10.1128/JVI.00560-06.
- 246 21. **Delmas, B., J. Gelfi, R. L'Haridon, L. K. Vogel, H. Sjostrom, O. Noren, and H. Laude.**
247 1992. Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV.
248 *Nature.* **357**:417-420. doi: 10.1038/357417a0.
- 249 22. **Reguera, J., D. Ordone, C. Santiago, L. Enjuanes, and J. M. Casasnovas.** 2011.
250 Antigenic modules in the N-terminal S1 region of the transmissible gastroenteritis virus spike
251 protein. *J. Gen. Virol.* **92**:1117-1126. doi: 10.1099/vir.0.027607-0; 10.1099/vir.0.027607-0.
- 252 23. **Reguera, J., C. Santiago, G. Mudgal, D. Ordone, L. Enjuanes, and J. M. Casasnovas.**
253 2012. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by
254 neutralizing antibodies. *PLoS Pathog.* **8**:e1002859. doi: 10.1371/journal.ppat.1002859;
255 10.1371/journal.ppat.1002859.
- 256 24. **Yeager, C. L., R. A. Ashmun, R. K. Williams, C. B. Cardellichio, L. H. Shapiro, A. T.**
257 **Look, and K. V. Holmes.** 1992. Human aminopeptidase N is a receptor for human coronavirus
258 229E. *Nature.* **357**:420-422. doi: 10.1038/357420a0.
- 259 25. **Godet, M., J. Grosclaude, B. Delmas, and H. Laude.** 1994. Major receptor-binding and
260 neutralization determinants are located within the same domain of the transmissible
261 gastroenteritis virus (coronavirus) spike protein. *J. Virol.* **68**:8008-8016.
- 262 26. **Jiang, S., L. Lu, L. Du, and A. K. Debnath.** 2013. A predicted receptor-binding and critical
263 neutralizing domain in S protein of the novel human coronavirus HCoV-EMC. *J. Infect.*
264 **66**:464-466. doi: 10.1016/j.jinf.2012.12.003; 10.1016/j.jinf.2012.12.003.

265

Fig.1



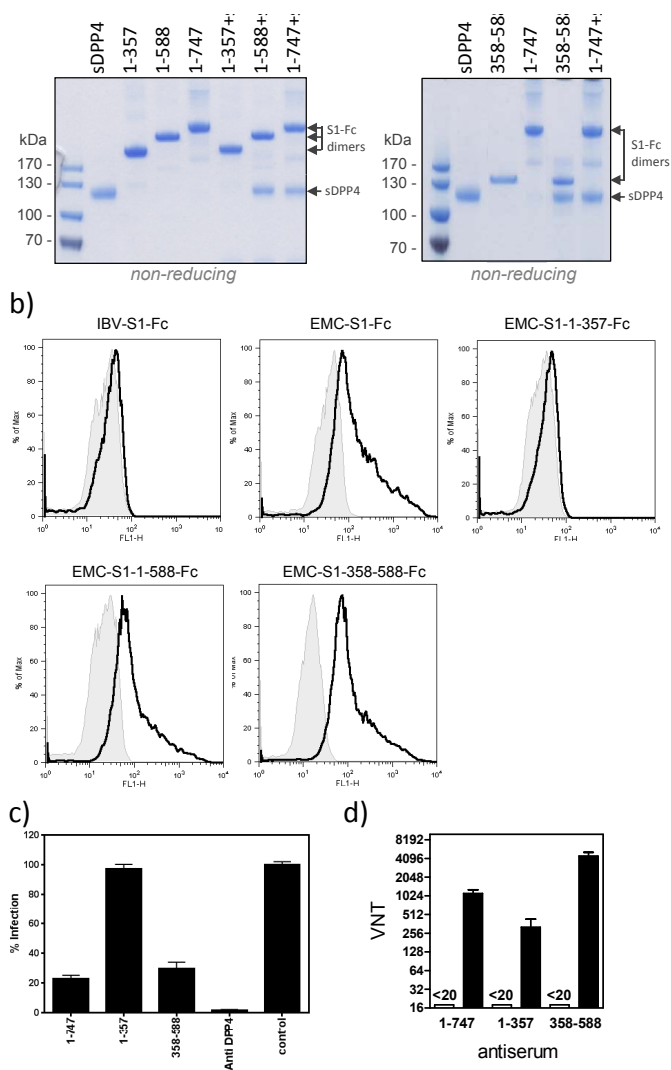


Fig.3

